

## **Big Data prédictif : jusqu'à quel point sommes-nous vraiment libres ?**

par **François Lainée, fondateur de Politics Angels**

**Scoring et Big Data. L'exploitation des données. Quelles sont les capacités prédictives des nouveaux outils numériques ? François Lainée reconnaît leur étonnant potentiel. Mais le Big Data peut aussi être vu comme une nouvelle boussole pour explorer les frontières d'une Terra Incognita : celle du libre arbitre de l'être humain.**

Dans un monde qui se numérise et se connecte de plus en plus, le Big Data est à la mode. Le Big Data c'est tout ce qu'il est ou sera possible de faire en utilisant intelligemment ce déluge de données.

Les entreprises, notamment celles dont les clients sont les particuliers, ont bien sûr commencé à explorer ces opportunités. Dans cette marche elles avancent par étapes, en commençant, souvent laborieusement, par rassembler et homogénéiser leurs données propres pour mieux connaître leur environnement. Puis elles cherchent quelles données externes elles peuvent y ajouter (Facebook ou Twitter sont alors des sources très prisées). Ensuite viennent l'analyse et la restitution de ces données, souvent en tableaux de bord adaptés aux métiers de l'entreprise.

Enfin, parfois, les entreprises s'attaquent au prédictif, en tentant de trouver dans les données des réponses à des questions sur les actions futures de leurs clients ou d'actifs **individuels**. Les questions abordées sont multiples : « Ce demandeur d'emprunt remboursera-t-il son prêt sur 10 ans ? », « Ce visiteur de mon site web va-t-il ou non finaliser son achat ? », « Quels sont les équipements de mon usine les plus à risque de tomber en panne ? », « Cet employé est-il à fort risque de démissionner ? ».

L'approche naturelle pour répondre à ces questions à partir des données porte le nom de scoring. Elle consiste, pour une question donnée, à construire une fonction mathématique qui produit un nombre, compris entre 0 et 1, et qui donne la probabilité que la réponse soit oui ou non (0 : l'événement va se matérialiser à coup sûr ; 1 : il n'y a aucune chance/ aucun risque). La pratique du scoring est ancienne. Il y a bien longtemps que les banques et assurances, notamment, construisent des scores de risques de leurs clients pour décider des octrois de prêts ou des souscriptions de police. Sous l'effet du Big Data, les approches de scoring évoluent elles aussi, gagnant en efficacité, et en universalité.

L'universalité, tout d'abord, est celle de la mesure de performance d'un score. Il existe maintenant une mesure reconnue, celle de la courbe ROC, valable en tout domaine, pour mesurer la qualité prédictive d'une règle de score. Sans entrer dans le détail, retenons ici que l'efficacité d'un score peut être caractérisée par un nombre, entre 0% et 100%, qui caractérise le côté totalement prédictif du score (valeur 100%) ou absolument non prédictif (valeur 0%).

L'efficacité est quant à elle liée à des avancées algorithmiques, notamment en matière d'apprentissage machine, un ensemble d'approches qui explorent des bases de données historiques pour « apprendre » ce qui fait un profil propre ou non à matérialiser l'évènement. Ces techniques statistiques utilisent la puissance des calculateurs pour explorer des quantités massives de données, tant en nombre d'individus qu'en nombre de paramètres descriptifs (âge, sexe, situation familiale, lieux de vie, métier, revenu pour un octroi de prêt, mais aussi mouvements détaillés du compte en banque depuis des années).

Et la puissance de ces approches commence à donner des résultats étonnants, portant des impacts économiques majeurs, et ouvrant des perspectives philosophiques.

Ainsi, dans le cas des scores d'octroi de prêts, le Big Data a déjà pu offrir des avancées qui peuvent sembler modestes, mais sont en fait majeures. Les scores traditionnels des banques sont souvent assez médiocres, avec des performances de moins de 25%. Avec les techniques d'apprentissage, cette performance peut augmenter de 5 à 10%, pour atteindre 35%. On est donc toujours loin d'être très prédictif, mais ce gain, traduit dans les comptes des banques en provisions pour risques réduites, peut devenir un levier de profit majeur. D'un point de vue prédictif, toutefois, le verdict est très clair : même en analysant ce qu'elle sait de vous, la banque ne sait que très imparfaitement dire si vous

rembourser votre prêt ou non. Cette question n'est pas déterministe, et les événements futurs, ceux que vous subirez ou choisirez d'engager (un divorce, un changement de travail, un accident, des choix de vie ou de consommation...) pourront influencer le résultat final.

Mais la prédiction est parfois presque parfaite. Dans le cas de cohortes de femmes atteintes ou non de cancer du sein, des scores efficaces à plus de 95% ont pu être construits, corrélant l'occurrence de la maladie à une combinaison complexe de paramètres démographiques et médicaux. Dans ces contextes, l'évolution d'un patient vers les zones de paramètres qui prédisent la maladie le place sur une voie certaine vers la maladie. Si cette évolution est elle-même inévitable (l'âge qui avance et des paramètres physiologiques « essentiels », que l'hygiène de vie ou la médecine ne savent pas influencer), on se trouve dans une vie prédite, d'où la liberté semble avoir disparu.

Ces exemples ne sont qu'une minuscule illustration des potentiels étonnants du Big Data pour tracer de nouvelles frontières à cette Terra Incognita qu'est notre libre arbitre. Data et scoring sont sur ces terrains de nouvelles boussoles. L'exploration commence. Et, en destination de ce nouveau voyage, j'écris ton nom : Liberté.

 [Télécharger le PDF de l'article](#)

<< [Retour au sommaire](#)

## PRES@ JE.COM

Une publication de l'Institut PRESAJE

**(Prospective, Recherche et Etudes Sociétales Appliquées à la Justice et à l'Economie)**

Siège social : 2 avenue Hoche 75008 Paris - Courrier : 30 rue Claude Lorrain 75016 Paris

Tél. 01 46 51 12 21 - E-mail : [contact@presaje.com](mailto:contact@presaje.com) - [www.presaje.com](http://www.presaje.com)

Directeur de la publication : Michel Rouger

Pour ne plus recevoir d'e-mails de la part de Presaje, [cliquez ici](#) >> [CONSULTER LES PRECEDENTS NUMEROS](#)